Semiparametric Model and Bayesian Analysis for Clustered Accelerated Life Testing Data

Rong Pan*[1], Yukiko Kozakai[2]

School of Computing, Informatics, and Decision Systems Engineering, Arizona State University

School of Mathematical and Statistical Sciences, Arizona State University

Tempe, Arizona, U.S.A

*[1]rong.pan@asu.edu; [2]yukiko.kozakai@asu.edu

Abstract

One of the difficulties in analyzing accelerated life testing data is the model-based failure probability prediction. Choosing an inferior model yields inaccurate predictions that can be exaggerated by extrapolation. Furthermore, testing data are often naturally clustered in groups, thus some modeling exibility must be granted to handle both the intra-cluster and inter-cluster variations. To address these problems, we discuss a data fitting strategy in this paper by developing a semiparametric model with random effects and the Bayesian piecewise exponential inference method. The proportional hazard model and Weibull accelerated failure time model are examined and compared. Our result suggests that the Bayesian piecewise exponential model with random effects outperforms other models.

Keywords

Frailty Model; Weibull Distribution; Proportional Hazard; Reliability Prediction; Goodness of Fit

Introduction

Clustered accelerated life testing (ALT) data are obtained when ALT experiments are conducted on different batches of materials or by different laboratories or experimenters. It is clear that the assumption of independent observations is no longer valid for clustered ALT data as the observations from same cluster are correlated. Such correlated data are often met in reliability engineering problems. For examples, Lindley and Singpurwalla (1986) discussed the case where the components of a system share the same use environment and a hash environment will encourage early failures on all components in one system; Pan and Crispin (2011) analyzed the clustered degradation data using a hierarchical modeling approach. Frailty, a statistical concept of the random effect of some latent variable, can be used to represent heterogeneity among clusters (Wienke 2003). Although the frailty model has provided a rich theoretical framework for modeling and analyzing survival data (see, Hougaard 2000; Duchateau and Janssen 2008), its advantage on analyzing ALT data was not well appreciated. Ma and Krings (2008) discussed some potential applications of the shared frailty model in reliability engineering and computer science engineering. A further generalized model, the correlated frailty model, was discussed in Stefanescu and Turnbull (2006).

The shared frailty model assumes that observations within a cluster share the same unknown effect, thus they are positively correlated. This model is suitable for analyzing clustered ALT data because the frailty variable can be used to quantify the variability between clusters. In order to propose a comprehensive modeling approach, this paper starts with an investigation of some popular ALT models, i.e., the Weibull accelerated failure time (AFT) model and the Cox proportional hazard (PH) model, and their shared frailty variants that consider random effects. The Bayesian inference is a powerful tool, as the Bayesian modeling has the flexibility in assigning heterogeneity by treating each model parameter as a random variable. Therefore, we develop a Bayesian piecewise exponential (PE) model as an extension of the shared frailty model. This model removes the assumption of failure time distribution and is thus considered as semiparametric. We use an example to demonstrate the usefulness of this model on modeling clustered ALT data and provide the techniques for assessing both the global and local fitness of models.

A Real Data Set

The following example is used to illustrate the type of problem that we are going to discuss. In manufacturing industry, ALT is often used to understand the causes behind failures and to predict the reliability of a product. Gerstle and Kuntz (1983) studied the reliability of pressure vessels wrapped by

different spools through an ALT experiment, in which each vessel was tested under a specific pressure level. Leon et al. (2007) proposed a random effect Weibull regression model to analyze the data. The main objective of their study was to predict the lifetime of pressure vessel while considering a random effect of spool.

Table 1 is the variable description of the pressure vessel data from Gerstle and Kunz (1983). The experiment was conducted at the U.S. Department of Energy Lawrence Livermore Laboratory. Each spool was used to wrap some pressure vessels, and then the failure times (hours) to burst the pressure vessels were recorded at four different pressure levels (MegaPascals). The data set consists of a total of 108 failure and censoring times observed from a random sample of eight spools. Figure 1 presents the box plots of failure times from each spool. One can see that these failure time distributions are skewed to the right. The entire data set is given in Appendix.

In this paper, we will explore the adaptability of the semiparametric PH model with a frailty term on modeling and analyzing this type of ALT data. Note that the underlying assumption is that the test units from the same spool are homogeneous.

Lifetime Regression Models

The accelerated failure time (AFT) model (Kalbfleisch and Prentice 1980) and the Cox's proportional hazard (PH) (Cox 1972) are two popular lifetime regression models in literature. The AFT model assumes a parametric distribution and a parametric function for the failure time and the effect of explanatory variables, respectively. On the other hand, the PH model has a parametric function for explaining the effect of explanatory variables only, without specifying the failure time distribution; thus, it is semiparametric. Random effects can be introduced in both models in either additive or multiplicative manner. The PH model with a multiplicative random effect term is often referred as the frailty model.

Shared Frailty Model

The concept of shared frailty was first introduced in Clayton (1978) for studying the familial tendency in disease incidence. When there is excessive variability among individual units in a experimental group or when there is an unmeasured factor that may affect the hazard function, we can multiply the hazard function of a PH regression model by a random

TABLE 1 VARIABLE DESCRIPTION OF PRESSURE VESSEL DATA

Variable Name	Type of Variable	Variable Description
failure time	continuous	time-to-failure until 41000 hours
censored	binary	1=observed, 0=right-censored
stress	continuous	4 levels: 23.4, 25.5, 27.6, 29.7 megapascals
spool	categorical	8 levels: Spool 1 to Spool 8

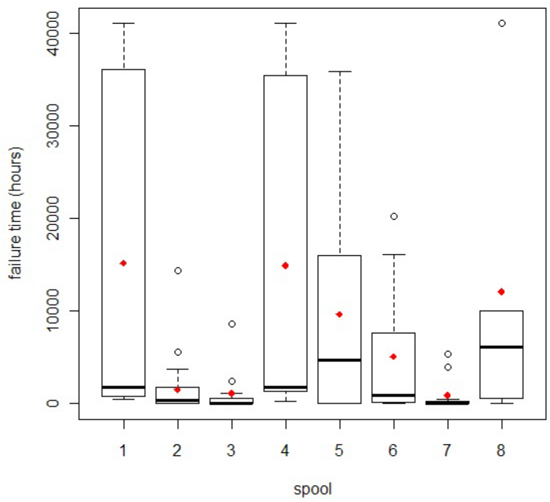

FIG. 1 BOX PLOTS OF THE TIMES TO FAILURE FROM EACH SPOOL

variable to account for the extra variability. Suppose we have k clusters and within each cluster all test units share the same cluster effect, then we have

$$\lambda_{ij}(t) = \lambda_0(t) z_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) , \qquad (1)$$

where $\lambda_{ij}(t)$ is the hazard function of the j^{th} test unit in the i^{th} cluster, $i = 1,2,...k$; $\lambda_0(t)$ is the baseline hazard function and its functional form needs not be specified in the PH model; z_i is the random effect term for the i^{th} cluster; $\mathbf{x}_{ij}$ and $\boldsymbol{\beta}$ are the explanatory variable vector (stress factors in ALT) and regression coefficient vector, respectively. The failure distribution function becomes

$$F_{ij}(t) = 1 - R_0(t)^{z_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta})} , \qquad (2)$$

where $R_0(t) = \exp(-\int_0^t \lambda_0(\tau) d\tau)$ is the baseline reliability function.

Weibull Regression Model

The Weibull regression model is an AFT model with Weibull lifetime distribution, i.e., the lifetime distribution function is

$$F_{ij}(t) = 1 - \exp\left[-\left(\frac{t}{\theta_{ij}} \right)^{\rho} \right], \tag{3}$$

where θ and ρ are the scale and shape parameter of Weibull distribution, respectively. The AFT model assumes that the logarithm of scale parameter is a linear function of explanatory variables. Adding a random effect term, w_i, it becomes

$$\log\theta_{ij} = \gamma_0 + \mathbf{x}_{ij}^T \boldsymbol{\gamma} + w_i. \tag{4}$$

Therefore, the effect of explanatory variables is reflected by changing the time scale of the lifetime of an individual test unit. Note that by logarithm transformation, the transformed failure time, $\log T$, has the smallest extreme value (SEV) distribution with a location parameter as $\mu_{ij} = \log\theta_{ij}$ and a scale parameter as $\sigma = 1/\rho$. It is easy to show that, after carefully specifying the baseline hazard function, a PH regression model may become identical to a Weibull regression model.

A fully parameterized frailty model includes a lifetime distribution model, a regression model that associates the distribution parameter with covariates, and a random effect distribution model or frailty distribution model that specifies the distribution of the frailty term. A semiparameterized model does not baseline hazard function, which can be treated as a nuisance term. To perform maximum likelihood estimation (MLE) for model parameters directly, one typically has to have a fully parameterized model. However, by treating the random effect variable as missing data Klein (1992) showed that an expectation-maximization (EM) algorithm, which iteratively taking expectation of full likelihood function and finding parameter estimates using the Cox's partial likelihood maximization method, can be implemented to estimate coefficients for covariates without specifying the baseline hazard function. A modified EM algorithm for the estimation in correlated frailty model was developed in Xue and Brookmeyer (1996). An alternative model, semiparametric AFT model with a multiplicative frailty term, was proposed in Pan (2001), in which the author developed an EM estimation method similar to the Klein's method. However, the EM algorithm may converge toward a local maximum instead of the global one (Shina and Day 1997). To circumvent this problem, a full Bayesian approach, which uses Markov Chain Monte Carlo (MCMC) sampling, can be applied. In addition, the MCMC method conveniently handles the computational problem that arise from the integration in high-dimensional space, so it enables us to analyze complex models such as the one for clustered ALT data.

Bayesian Inference

In the following section we describe the Bayesian inference method for clustered ALT data using both the AFT and PH models. There are two main reasons for choosing Bayesian inference. Firstly, industrial ALT experiments are typically very expensive and sample size is usually limited. Therefore, if one can integrate the previous knowledge on the lifetime of the tested product or similar products and/or expert opinions into the data analysis process, the estimation and prediction precision can be greatly improved. Secondly, in engineering it is generally preferable to communicate the failure prediction of a product in a probabilistic form. A single point estimation does not provide adequate information for engineering decision, especially when product failures are to be avoided.

Weibull AFT Bayesian Inference Method

We assume that failure times follow a Weibull distribution and that the frailty terms z_i are either gamma or lognormal distributed. In this model, Y_{ij}, the failure time variable of the j^{th} test unit in the i^{th} cluster, has the failure density function as

$$f(y_{ij}) = \frac{\rho}{\theta_{ij}} \left(\frac{t}{\theta_{ij}} \right)^{\rho-1} \exp\left[-\left(\frac{t}{\theta_{ij}} \right)^{\rho-1} \right],$$

where $\theta_{ij} = z_i \exp(\gamma_0 + \mathbf{x}_{ij}^T \boldsymbol{\gamma})$ as by Equation (4), and $z_i = \exp(w_i)$, which is the random covariate for the i^{th} cluster. Each random covariate is assumed to be independent and identically distributed with either gamma or lognormal prior, such as,

$$z_i \sim \Gamma(\eta, \eta)$$

or

$$\log z_i \sim N(0, \sigma_z^2),$$

so that the mean of z_i is 1 and η or σ_z^2 controls the disperse of the prior distribution. To complete the model specification, we set the prior distributions for ρ and $\boldsymbol{\beta}$ to be

$$\rho \sim N(0,\sigma_\rho^2)$$

and

$$\boldsymbol{\beta} \sim N_p(\mathbf{0},\sigma^2\mathbf{I}_p).$$

Using Equation (3), the likelihood function of this model is

$$L(\rho,\mathbf{z}) = \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left(\frac{\rho}{\theta_{ij}}\left(\frac{y_{ij}}{\theta_{ij}}\right)^{\rho-1}\right)^{\delta_{ij}} \exp\left(-\left(\frac{y_{ij}}{\theta_{ij}}\right)^{\rho}\right), \quad (5)$$

where δ_{ij} is the censoring indicator for the j^{th} unit in the i^{th} cluster. The posterior density $f(\rho,\boldsymbol{\beta},\mathbf{z}\,|\,D)$ becomes

$$f(\rho,\boldsymbol{\beta},\mathbf{z}\,|\,D) \propto L(\rho,\boldsymbol{\beta},\mathbf{z})\, f(\rho)\, f(\boldsymbol{\beta})\prod_{i=1}^{k} f(z_i).$$

To obtain posterior samples, we utilize the hybrid Metropolis-Gibbs sampling algorithm.

Semiparametric Bayesian Inference Method

Among various Bayesian approaches, the piecewise exponential Bayesian inference method provides an extremely flexible framework for modeling reliability data. This approach can be roughly described as a Bayesian counterpart to the frequentist approach of the Cox PH model. For the detail of piecewise exponential hazard model, we refer readers to Ibrahim et al. (2001).

In this model the time axis is divided into M pre-specified intervals $I_m = (s_{m-1}, s_m]$, $m = 1,\ldots,M$ so that $0 = s_0 < s_1 < \ldots < s_M$ with s_M being the last failure time or the censoring time. The model then assumes constant baseline hazard rate in each interval, i.e.,

$$\lambda_0(t) = \lambda_m \text{ for } t \in I_m.$$

Similar to the PH model, the hazard function $\lambda_{ij}(t)$ for the j^{th} individual in the i^{th} cluster is assumed to be proportional to the baseline hazard function $\lambda_0(t)$, i.e.,

$$\begin{aligned}\lambda_{ij}(t) &= z_i \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta})\lambda_0(t) \\ &= z_i \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta})\lambda_m I(t \in I_m)\end{aligned} \quad (6)$$

where I denotes the indicator function for $t \in I_m$.

Now let y_{ij} be the failure or censoring time of the j^{th} test unit in the i^{th} cluster. Define

$$\Delta_{ijm} = \min\{y_{ij}, s_m\} - s_{m-1}. \quad (7)$$

That is, Δ_{ijm} is the time that this test unit has spent in the interval I_m. Denote g_{ij} to be an index such that $y_{ij} \in (s_{g_{ij}}, s_{g_{ij}+1}] = I_{g_{ij}+1}$. The data set becomes $D = \{\Delta_{ijm}, g_{ij}, \mathbf{x}_{ij}, i = 1,\ldots,k,\ j = 1,\ldots,n_i, m = 1,\ldots,M\}$. The complete likelihood function becomes

$$\begin{aligned}L(\beta,\lambda,z\,|\,D) \propto &\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left(\prod_{m=1}^{g_{ij}} z_i\lambda_m\Delta_{ijm}e^{\mathbf{x}_{ij}^T\boldsymbol{\beta}}\right) \\ &\times\prod_{i=1}^{k}\prod_{j=1}^{n_i}(z_i\lambda_{g_{ij}+1}e^{\mathbf{x}_{ij}^T\boldsymbol{\beta}})^{\delta_{ij}} \exp(-z_i\lambda_{g_{ij}+1}\Delta_{ijm}e^{\mathbf{x}_{ij}^T\boldsymbol{\beta}})\end{aligned} \quad (8)$$

In this model the baseline hazard function assumes the form of step function much like the obtained baseline hazard using the PH model. The distinction between this model and the frequentist model is that the baseline hazard rate m is a random variable. Same as the frailty variable, the common choice of prior distribution for the piecewise constant baseline hazard rate is gamma or lognormal distribution. Using the gamma prior distribution, we assume

$$\lambda_m\,|\,(\lambda_{m-1},\ldots\lambda_1) \sim \Gamma(\alpha_m, \alpha_m/\lambda_{m-1}),\ m = 1,\ldots,M,$$

where α is a smoothing parameter. Using the lognormal prior distribution, we let $\lambda_m = \exp(\xi_m)$ and

$$\xi_m\,|\,(\xi_{m-1},\ldots,\xi_1) \sim N(\xi_{m-1},\sigma_\xi^2).$$

To complete the specification, we set the prior distribution for z_i and $\boldsymbol{\beta}$ by

$$z_i \sim \Gamma(\eta,\eta)$$

or

$$\log z_i \sim N(0,\sigma_z^2)$$

and

$$\boldsymbol{\beta} \sim N_p(\mathbf{0},\sigma^2\mathbf{I}_p).$$

Using each of the prior densities, the posterior density $f(\lambda, \boldsymbol{\beta}, \mathbf{z} \mid D)$ is proportional to

$$f(\lambda, \boldsymbol{\beta}, \mathbf{z} \mid D) \propto L(\boldsymbol{\beta}, \lambda, \mathbf{z} \mid D)\, f(\lambda)\, f(\boldsymbol{\beta})\, f(\mathbf{z}),$$

where $f(\lambda)$, $f(\boldsymbol{\beta})$, and $f(\mathbf{z})$ represent the respective prior and conditional prior density functions of parameters and frailty variables.

Ibrahim et al. (2001) derived the form of the conditional posterior densities with the gamma prior distribution for frailty terms so that a Gibbs sampling approach is possible. However, with other prior distributions the analytical forms of full condition posterior distributions cannot be obtained; thus we use an adaptive acceptance method, the Metropolis-Gibbs sampling algorithm, to solve this problem. In addition, this method overcomes the shortcoming of the Gibbs sampling when the log-concavity of conditional posterior distribution is not guaranteed. Our model is implemented in WinBUGS, a popular MCMC software for Bayesian inference, and it will be described in the next section.

Analyzing the Industrial Example

In this section we apply Bayesian data analysis on the ALT data from the pressure vessel testing described in Section 1.1.

Analysis of Parametric Models

The AFT model is the most common ALT model for reliability testing in literature. However, the AFT model assumes a specific failure time distribution, which must be verified. Table 2 displays the loglikelihood scores for the exponential, Weibull and lognormal models. The results indicate that the Weibull AFT model provides the best fit to this data set, but a simpler model, the exponential AFT model, comes close to it. A useful plot is constructed by taking the logarithm of the Nelson-Aalen estimator of the cumulative hazard rate versus the logarithm of time.

TABLE 2 TEST FOR PARAMETRIC DISTRIBUTION ASSUMPTIONS

Distribution	Max log-likelihood	df	AIC
Exponential	-740.87	9	1499.74
Weibull	-737.07	10	1491.14
Lognormal	-777.92	10	1575.84

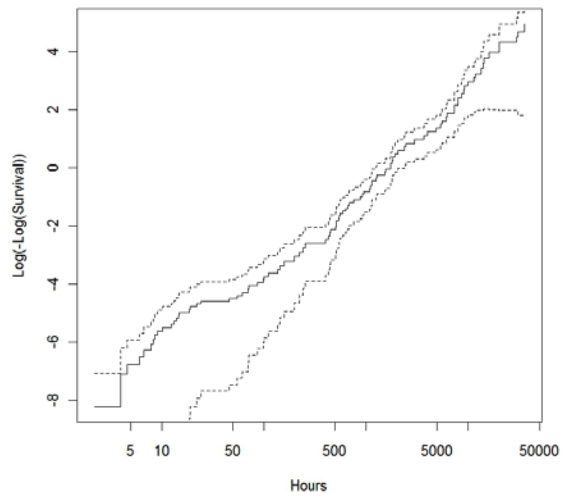

FIG. 2 CUMULATIVE HAZARD FUNCTION

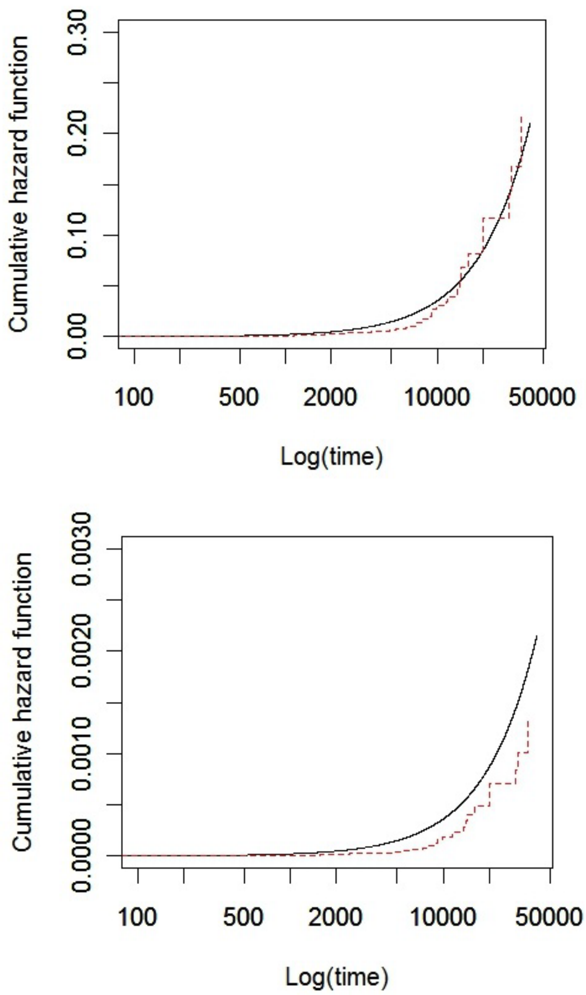

FIG. 3 COMPARISON OF THE CUMULATIVE HAZARD FUNCTION OF WEIBULL MODEL (BLACK SOLID LINE) AND PH MODEL (RED BROKEN LINE) WITH THE STRESS LEVEL AT 23.4MPA (TOP) AND 20MPA (BOTTOM)

Under the assumption that the failure time have a Weibull distribution, the plot should be a straight line with slope ρ. The plot in Figure 2 is shown for the

case when the stress level is set at the its mean value (log(stress)=3.181). It seems to be a straight line, thus one might think that the Weibull model fit fairly well. This is confirmed by plotting the cumulative hazard function of the Weibull model against the cumulative hazard function of the PH model for Spool 1 at the lowest tested stress level 23.4MPa, as shown in the left panel of Figure 3. However, when we extrapolate the cumulative hazard function to the 20MPa stress level, the right panel of Figure 3 shows that the fitted Weibull distribution has a constantly higher cumulative hazard function than the PH model. This indicates that the apparent good fitting of Weibull distribution may not always hold when model extrapolation has to be made. Unfortunately, for most ALT experiments the data extrapolation is unavoidable in order to draw inference to the product's failure rate at its use condition. Therefore, we ought to exercise a great deal of caution on evaluating competing models.

Testing Proportionality Assumption

The difficulty of fitting a parametric model can be resolved by using a semiparametric model such as the Cox's PH model. A critical model assumption of the PH model is the proportionality hazard assumption, which stipulates that covariate levels affect the hazard function uniformly over time by multiplying by a constant, $\exp(\mathbf{x}^T\boldsymbol{\beta})$. Under this assumption, the Kaplan-Meier estimators for the survival function for each level of the covariate should appear to be nearly parallel curves, i.e., the graph of $-\log(-\log \hat{R}(t))$ versus $\log t$ results in parallel lines.

We plot the Kaplan-Meier estimators for the covariate (logarithm of stress) in Figure 4, which shows that the curves of the tested stress levels are roughly parallel, except of the curve of the highest stress level. It could be caused by spool effects, as the spools used in the different stress tests are different in general. Therefore, we further employ the scaled Schoenfeld residuals to test the proportionality assumption. This method works by adding a time-dependent covariate $x \times g(t)$ to the PH model for some known function $g(t)$, then a plot of the scaled residual versus time can provide some visual evidence of nonproportionality (see Therneau and Grambsch (2000) for details). The logarithm of time $g(t) = \log t$ is often chosen for the data having a long-tailed lifetime distribution. We performed the scaled Schoenfeld residual plots for the fixed covariate, $\log(stress)$, and it appears to be horizonal lines; thus there is no serious concern of the proportionality assumption.

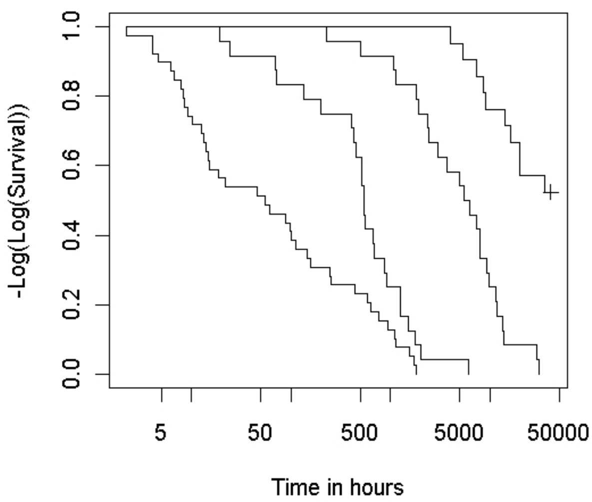

FIG. 4 KAPLAN-MEIER ESTIMATOR FOR LOG(STRESS), FROM THE LEFT, 29.7, 27.6, 25.5, 23.4MPA, RESPECTIVELY

TABLE 3 MODEL COMPARISON IN PRELIMINARY STUDY

Model	Max. Log-likelihood	AIC
Weibull AFT with gamma frailty	-737.24	1487.94
Weibull AFT with lognormal frailty	-737.18	1487.87
PH with gamma frailty	-262.29	526.58
PH with log3normal frailty	-262.17	526.33

Our preliminary model fitting is performed on both the PH and AFT models. Gamma distribution and lognormal distribution are used for modeling the random effects of spool, and the maximum likelihood method (the EM method in the case of PH model) is employed to estimate parameters. A standard method of assessing the overall goodness of fit between competing models is based on the likelihood value and the Akaike's information criterion (AIC). Table 3 gives these measures on respective models. Comparing their maximum log-likelihood values, one can see that the PH frailty models have much larger score. This is expected as the semiparametric model would not force the hazard function to strictly follow the Weibull distribution assumption in any time period, so it fits the data better globally. A similar result can be seen from the AIC criterion (smaller the better). Note that the AIC statistic compromises between model fitting and model complexity. The fact that the AIC statistic is smaller in this case shows that the semiparametric PH model increases the fit to the

data sufficiently enough to offset the additional model complexity needed by the PH model. Overall, the PH frailty model with lognormal frailty distribution has the best goodness of fit.

Bayesian Analysis of the PH Frailty Model

We develop a piecewise exponential model, as described in the previous section, for the pressure vessel data. Note that, strictly speaking, this model is not nonparametric in distribution modeling, but it approximates the PH model well when its intervals are carefully chosen. Furthermore, it reduces the number of baseline hazard parameters significantly and thus would not suffer the loss of inference precision as the PH model would. Past studies suggested that there should be at least one failure observation in each interval. We divide the whole testing period into 10 intervals and there are roughly 10 failure observations in each interval. The data and the time intervals are given in Appendix.

Based on the relationship of exponential failure time distribution and Poisson failure count distribution, we model the failure indicator of each test unit (i.e., the j^{th} unit in the i^{th} cluster) in each time interval by a Poisson distribution with its mean value being the product of failure rate and time interval. These failure indicators are binary variables and they are independent to each other. Therefore, the contribution of each failure indicator to the total log-likelihood function, when the test unit failed in the m^{th} interval, or survived the m^{th} interval, or is cencored before the m^{th} interval, is, respectively,

$$ll_{ijm} = \begin{cases} \mu_{ijm} + \log \mu_{ijm} \\ \mu_{ijm} \\ 0 \end{cases} \tag{9}$$

where $\mu_{ijm} = \lambda_{ijm} \times \Delta_{ij}$, and λ_{ijm} and Δ_{ijm} are defined as in Equations (6) and (7), respectively. The complete likelihood function is as $L = \exp(\sum_i \sum_j \sum_m ll_{ijm})$.

We use the lognormal prior distribution for the frailty term (spool). The MCMC is implemented in WinBUGS. A total of 20000 MCMC samples are taken. Both Gelman-Rubin and Raftery-Lewis criteria are used to confirm the convergence of the MCMC chain. The first 10000 samples are discarded as burn-ins and the second 10000 samples are used to calculate the posterior estimation of the model parameter of interest.

For example, one may be interested in the effect of each cluster on the hazard function. Figure 5 shows the predicted effect of each spool, i.e., z_i in Equation (6). As the medians of z_2, z_3, z_6 and z_7 are larger than 1 and the others are smaller than 1, the first group of spools has shorter expected lifetime than the second group. For instance, the lifetime of a unit from Spool 1 is expected to be 12.4 (i.e., $e^{\hat{z}_2 - \hat{z}_1}$) times longer than the lifetime of a unit from Spool 2.

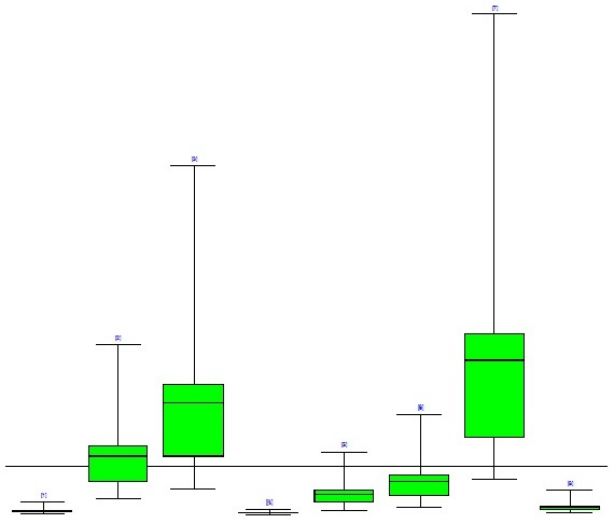

FIG. 5 BOX PLOTS OF INDIVIDUAL SPOOL EFFECTS WITH THE MIDDLE HORIZONTAL LINE AS THE GRAND AVERAGE

TABLE 4 POSTERIOR ESTIMATION OF z_i

Spool	Median	95% Interval
z_1	0.1942	(0.05698, 0.7046)
z_2	2.712	(0.8879, 9.863)
z_3	4.941	(1.499, 20.23)
z_4	0.09229	(0.2648, 0.3309)
z_5	0.9298	(0.261, 3.637)
z_6	1.533	(0.4636, 5.873)
z_7	6.727	(2.051, 29.07)
z_8	0.3768	(0.1118, 1.392)

To compare the piecewise exponential model with other models, we implement the Bayesian data analysis using the Weibull AFT model as described in Leon et al. (2007). Our aim here is to compare both the global and local goodness of fit of the two models. As aforementioned, model extrapolation is often applied on ALT model and a small difference in modeling can

lead to a large discrepancy in prediction, so we need to carefully examine the merit of each candidate model.

Table 5 presents the deviance information criteria (DIC) of both models. The DIC is the equivalent to AIC in Bayesian analysis. It combines the evaluation of the likelihood function based on posterior samples of model parameters and the consideration of model complexity (Spiegelhalter et al. 2002). A much smaller DIC value of the piecewise exponential PH model than the Weibull AFT model indicates that the former model fits the overall data better. Assuming that our main interest is the failure probability of a vessel at 1000 hours, we further explore the fitness of the two models around this local time by cross validation. Nine failures have happened between 700 and 1200 hours in the data set. We jackknife these 9 failure time observations one at a time, construct both models, and compare the model prediction of the removed failure time with its true value. Mean square error (MSE) is calculated for each of the jackknifed failure data over 10000 posterior samples for each model. They are listed in Table 6. The PH model has a smaller average MSE value than the AFT model, which indicates that it provides a better fit the local time of 1000 hours.

TABLE 5 COMPARISON OF GLOBAL DATA FITTING OF TWO MODELS. $\overline{D}$ IS THE POSTERIOR MEAN DEVIANCE; $\hat{D}$ IS THE DEVIANCE OF POSTERIOR MEANS; PD MEASURES MODEL COMPLEXITY, $\mathrm{pD} = \overline{D} - \hat{D}$; AND $\mathrm{DIC} = \overline{D} + \mathrm{pD}$

Model	$\overline{D}$	$\hat{D}$	pD	DIC
Piecewise exponential PH model	488.17	471.80	16.37	504.54
Weibull AFT model	1481.73	1472.43	9.30	1491.04

TABLE 6 CROSS-VALIDATIONS OF LOCAL DATA FITTING OF THE TWO MODELS

Obs.	Spool	Stress	True value	MSE by PH model ($\times 10^5$)	MSE by AFT model ($\times 10^5$)
1	1	29.7	755.2	1.78	1.85
2	1	29.7	1108.2	6.19	5.97
3	4	29.7	1148.5	3.41	2.84
4	4	27.6	876.7	84.68	107.14
5	1	27.6	930.4	15.53	18.32
6	6	27.6	1254.9	7.79	7.70
7	4	27.6	1275.6	65.44	89.66
8	3	25.5	1087.7	1.18	1.04
9	2	25.5	1134.3	2.36	6.60
Ave.				20.93	26.79

Given that the failure probability at 1000 hours is of interest, we compare the results from the two model at different stress levels. Table 7 lists the point estimation (posterior median) and 95% credible interval of failure probability under 29.7MPa (the highest test stress level), 23.4MPa (the lowest test stress level) and 20MPa (the hypothesized use stress level).

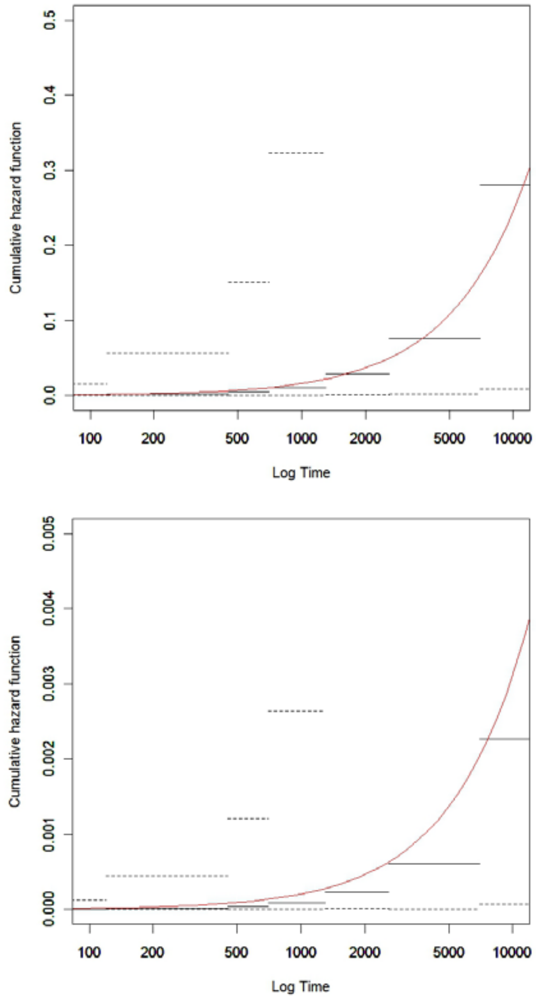

FIG. 5 PREDICTED CUMULATIVE HAZARD FUNCTION AT THE STRESS LEVELS OF 23.4MPA (TOP) AND 20MPA (BOTTOM) FROM THE TWO MODELS

TABLE 7 PREDICTIONS OF FAILURE PROBABILITY AT THREE STRESS LEVELS BY THE TWO MODELS

Model	Failure prob. at 1000 hours stress=29.7MPa	Failure prob. at 1000 hours stress=23.4MPa	Failure prob. at 1000 hours stress=20MPa
Piecewise exp. PH model	1 (0.3584, 1)	0.01 (2.932e-4, 0.2762)	8.098e-5 (2.285e-6, 2.632e-3)
Weibull AFT model	1 (0.2305, 1)	0.01514 (3.753e-4, 0.4189)	2.034e-4 (3.484e-6, 8.024e-3)

One can see that the discrepancy between the two models grows as we apply extrapolation from the test stress level to the use stress level. At the 23.4MPa stress level, the AFT model predicts the failure probability 1.5 times larger than the predicted value of the PH model, while it is 2.5 times larger at the 20MPa stress level. Note that the credible intervals from the PH model are smaller than those from the AFT model, suggesting a better global fit of the former model. However, the PH model cannot estimate the hazard rate beyond the testing period, which is the limitation of semiparametric model; one has to depend on a full parametric model for such exploration.

Effects of Sample Size and Test Suration

To reveal the effects of sample size and test duration on the failure probability prediction and to compare the performance of the two models – the semiparametric PH model and the Weibull AFT model, we conduct a simulation study using the model derived from this industrial example. As aforementioned, the results from ALT experiments are often extrapolated to the use stress condition to make the prediction of a failure quantity of interest. Again, we are interested in the failure probability of vessel under 20MPa at 1000 hours. In order for a parameterized model to perform well, it typically requires a large number of failure observations, which could be a difficult condition for ALT due to time and cost. We have seen from this example that there are four stress levels and under each stress level there are less than 20 test units. Due to censoring, failure time may not be observable from every single test unit. The limited sample size and incomplete data hinder the accuracy of any model fitting, and the model-based extrapolation can exaggerate model bias and lead to error-prone prediction.

In the simulation, data are generated from a Weibull distribution with shape and scale parameters to be 1.266 and $\exp(84.47 - 23.03 \times \log(\text{stress level})$, respectively. This corresponds to the accelerated failure time model of Spool 1. We vary the number of test units at each stress level (29.7MPa, 27.6MPa, 25.5MPa, 23.4MPa) from 30 to 10 and vary the censoring time from 40000 hours to 10000 hours. In each simulation run, a data set of failure or censoring times at four stress levels are generated. They are used to fit the two competing models – the piecewise exponential PH model and the Weibull AFT model. Based on the fitted model, we predict the failure probability at 1000 hours+ under the use stress of 20MPa. Given the true probability value from the assumed Weibull distribution, we can compute the difference between the model-based prediction value and the true value. There are 100 simulation runs for each case of sample size and test duration. The simulation is conducted in *R* by calling a function called "R2WinBUGS" to activate the WinBUGS codes to perform Bayesian analysis. The *R* and WinBUGS programs are available from the first author upon request. Two quantities are calculated: one is the root mean square error (RMSE) of the posterior median estimator of failure probability and the other one is the percentage of the true probability value falling into the 90% credible interval of failure probability. Tables 8 and 9 list their values for both models.

TABLE 8 EFFECT OF SAMPLE SIZE (CENSORING TIME = 41000 HOURS)

number of test units at each stress level	RMSE ($\times 10^{-4}$)		Percentage of Coverage	
	Piecewise Exp. PH	Weibull AFT	Piecewise Exp. PH	Weibull AFT
30	0.70	1.69	48%	20%
25	0.84	2.00	51%	25%
20	1.64	2.43	45%	28%
15	2.19	2.31	48%	34%
10	4.39	3.70	49%	51%

TABLE 9 EFFECT OF TEST DURATION (NUMBER OF TEST UNITS AT EACH STRESS LEVEL = 20)

censoring time (hours)	RMSE ($\times 10^{-4}$)		Percentage of Coverage	
	Piecewise Exp. PH	Weibull AFT	Piecewise Exp. PH	Weibull AFT
40000	1.47	2.18	50%	28%
30000	1.35	3.19	49%	17%
20000	1.67	6.74	49%	4%
10000	1.92	20.1	53%	0%

Since the RMSE is a measure of the deviation of the predicted value of failure probability to the true

probability value for the assumed failure time distribution, this measure is the smaller the better. The percentage of coverage measures the chance of the true failure probability value being enclosed by the 90% credible interval, so this measure is the larger the better. Note that the true failure time distribution model assumed in this study is Weibull distribution; however, the prediction results obtained from the Weibull AFT Bayesian analysis are uniformly worse than those from the piecewise exponential PH Bayesian analysis. Due to time censoring, we cannot observe failures on some test units, particularly at low stress levels. This causes the bias in the estimated Weibull model and, consequentially, large errors when using this model to predict the failure probability at the use stress level. This problem is somewhat alleviated when the semiparametric PH model is in use. From Table 8 one can see that when the number of test units is reduced both models perform worse in terms of their RMSEs, but in general the piecewise exponential PH model has fewer errors than the Weibull AFT model. From Table 9 one can see that the effect of censoring on prediction is profound when the Weibull AFT model is applied. As the test duration is shortened, more failure times are censored, particularly at low stress levels. The parametric Weibull model to be built will heavily rely on the failure data at high stress levels, which causes the over-prediction of failure probability when the model is extrapolated to the use stress level. Comparatively, the piecewise exponential model is more robust, as its performance degrades only slightly even when there are many failure times being censored.

Conclusion

In this paper, we discuss different lifetime regression models for clustered ALT data. The random effect of cluster variable is introduced so that we can draw inference on the whole population rather than a specific cluster. We demonstrate the flexibility of semiparametric approach and develop the Bayesian analysis of piecewise exponential model. It is important to emphasize that, in the process of developing Bayesian analysis for any model, it is critical to check model assumptions so that the comparison of global and local goodness of fit of competing models can be valid. As the model-based extrapolation is often employed when analyzing ALT data, a small variation in models can lead to a sizable difference in final conclusions. Reliability engineer should be prudent to choose the best model to insure the quality of ALT data analysis.

REFERENCES

Clayton, D. G. “A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence.” Biometrika 65(1978): 141-151.

Cox, D. R. “Regression Models and Life Tables.” Journal of the Royal Statistical Society, Series B 20(1972): 187-220.

D. J. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. “Bayesian Measures of Model Complexity and Fit.” Journal of the Royal Statistical Society, Series B 64(2002): 583-639.

Duchateau, L. and Janssen, P. The Frailty Model, Springer-Verlag New York, LLC., 2008.

Gerstle, E. P. and Kunz, S. C. “Prediction of Long-Term Failure in Kevlar 49 Composites.” American Society for Testing and Materials 813(1983): 263-292.

Hougaard, P. Analysis of Multivariate Survival Data, Springer-Verlag New York, Inc., 2000.

Ibrahim, J. G., Chen, M.-H., Sinha, D. Bayesian Survival Analysis, Springer-Verlag New York, Inc., 2001.

Kalbfleisch, J. D. and Prentice, R. L. The Statistical Analysis of Failure Time Data, John Wiley & Sons, New York, 1980.

Klein. J. P. “Semiparametric Estimation of Random Effects using the Cox Model Based on the EM Algorithm.” Biometrics 48(1992): 795-806.

Leon, R. V., Ramachandran, R., Ashby, A. J., Thyagarajan, J. “Bayesian Modeling of Accelerated Life Tests with Random Effects.” Journal of Quality Technology 39(2007): 3-16.

Lindley, D. V. and Singpurwalla, N. D. “Multivariate Distributions for the Life Lengths of Components of A System Sharing A Common Environment.” Journal of Applied Probability 23(1986): 418-431.

Ma, Z. and Krings, A. W. “Multivariate Survival Analysis (I): Shared Frailty Approaches to Reliability and Dependence Modeling.” in Proceedings of IEEE

Aerospace Conference, Big Sky, MT, March 2008.

McGilchrist, C. A. and Aisbett, C. W. "Regression with Frailty in Survival Analysis." Biometrics 47(1991): 461-466.

Pan, R. and Crispin, T. "A Hierarchical Modeling Approach to Accelerated Degradation Testing Data Analysis - A Case Study." Quality and Reliability Engineering, International 27(2011): 229-237.

Pan, W. "Using Frailties in the Accelerated Failure Time Model." Lifetime Data Analysis 7(2001): 55-64.

Sinha, D. and Dey, D. K. "Semiparametric Bayesian Analysis of Survival Data." Journal of the American Statistical Association 92(1997): 1195-1212.

Stefanescu, C. and Turnbull, B. W. "Multivariate Frailty Models for Exchangeable Survival Data with Covariates." Technometrics 48(2006): 411-417.

Therneau, T. M. and Grambsch, P. M. Modeling Survival Data: Extending the Cox Model, Springer Science+Bussiness Media, LLC., 2000.

Wienke, A. "Frailty Models." Working Paper of the Max Planck Institute for Demographic Research, Germany, 2003.

Xue, X. and Brookmeyer, R. "Bivariate Frailty Model for the Analysis of Multivariate Survival Time." Lifetime Data Analysis 2(1996): 277-289.

Yashin, Anatoli I., Vaupel, James W., Iachine, Ivan A. "Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data." Mathematical Population Studies 5(1995): 145-159.

APPENDIX

TABLE 10 DATA AND TIME INTERVALS USED IN THE PIECEWISE EXPONENTIAL MODEL

Interval	Observation (spool, pressure, failure time)
(0, 10]	(2, 29.7, 2.2), (7, 29.7, 4.0), (7, 29.7, 4.0), (7, 29.7, 4.6), (7, 29.7, 6.1), (6, 29.7, 6.7), (7, 29.7, 7.9), (5, 29.7, 8.3), (2, 29.7, 8.5), (2, 29.7, 9.1)
(10, 30]	(2, 29.7, 10.2), (3, 29.7, 12.5), (5, 29.7, 13.3), (7, 29.7, 14.0), (3, 29.7, 14.6), (6, 29.7, 15.0), (3, 29.7, 18.7), (3, 27.6, 19.1), (2, 29.7, 22.1), (3, 27.6, 24.3)
(30, 120]	(7, 29.7, 45.9), (2, 29.7, 55.4), (7, 29.7, 61.2), (3, 27.6, 69.8), (2, 27.6, 71.2), (5, 29.7, 87.5), (1, 29.7, 92.2), (8, 29.7, 98.2), (3, 29.7, 101.0), (2, 29.7, 111.4)
(120, 450]	(3, 27.6, 136.0), (6, 29.7, 144.0), (2, 29.7, 158.7), (2, 27.6, 199.1), (6, 25.5, 225.2), (5, 29.7, 243.9), (4, 29.7, 254.1), (2, 27.6, 403.7), (2, 27.6, 432.2), (1, 29.7, 444.4)
(450, 700]	(1, 27.6, 453.4), (7, 25.5, 503.6), (2, 27.6, 514.1), (6, 27.6, 514.2), (6, 27.6, 541.6), (2, 27.6, 544.9), (8,27.6, 554.2), (8, 29.7, 590.4), (8, 29.7, 638.2), (1, 27.6, 664.5), (2, 27.6, 694.1)
(700, 1300]	(1, 29.7, 755.2), (4, 27.6, 876.7), (1, 27.6, 930.4), (3, 25.5, 1087.7), (1, 29.7, 1108.2), (2, 25.5, 1134.3), (4, 29.7, 1148.5), (6, 27.6, 1254.9), (4, 27.6, 1275.6)
(1300, 2600]	(4, 27.6, 1536.8), (4, 29.7, 1569.3), (4, 29.7, 1750.6), (1, 27.6, 1755.5), (4, 29.7, 1802.1), (2, 25.5, 1824.3), (2, 25.5, 1920.1), (8, 27.6, 2046.2), (2, 25.5, 2383.0), (3, 25.5, 2442.5)
(2600, 7000]	(8, 25.5, 2974.6), (2, 25.5, 3708.9), (7, 23.4, 4000.0), (8, 25.5, 4908.9), (7, 23.4, 5376.0), (2, 25.5, 5556.0), (4, 27.6, 6177.5), (6, 25.5, 6271.1)
(7000, 12000]	(6, 23.4, 7320.0), (8, 25.5, 7332.0), (8, 25.5, 7918.7), (6, 25.5, 7996.0), (3, 23.4, 8616.0), (5, 23.4, 9120.0), (8, 25.5, 9240.3), (8, 25.5, 9973.0), (1, 25.5, 11487.3), (5, 25.5, 11727.1)
(12000, 41000]	(4, 25.5, 13501.3), (1, 25.5, 14032.0), (4, 25.5, 29808.0), (1, 25.5, 31008.0), (2, 23.4, 14400.0), (6, 23.4, 16104.0), (5, 23.4, 20231.0), (6, 23.4, 20233.0), (5, 23.4, 35880.0), (1, 23.4, 41000+), (1, 23.4, 41000+), (1, 23.4, 41000+), (1, 23.4, 41000+), (4, 23.4, 41000+), (4, 23.4, 41000+), (4, 23.4, 41000+), (4, 23.4, 41000+), (8, 23.4, 41000+), (8, 23.4, 41000+), (8, 23.4, 41000+)